Collaboratory Research and Development

New tools, facilities, and technologies are rapidly changing the practice and pace of modern science. High-throughput experiments and simulations are producing data that support new informatic sciences. Collaborations frequently involve large communities and are increasingly more dependent on shared data and tools published by others. Researchers are finding that they need to participate in multiple communities, often producing data or tools in one collaboration, while consuming information from another.

Collaboratory for Multi-scale Chemical Science (CMCS)

The Collaboratory for Multi-scale Chemical Science (CMCS) project [1] is incorporating advances in informatics, semantic web, collaboratory, and grid communities to facilitate collaboration among and within interdisciplinary combustion-science communities. The result is a first-generation combustion knowledge grid, where communities share data and analysis tools as they create verified, documented data sets and reference data so that it can be discovered; where new applications are accessible to wider audiences; and where new research directions have been identified as a result of community interaction.

Knowledge Environment for Collaborative Science (KnECS)

The infrastructure for CMCS is a domain-independent architecture for supporting collaborative science. This infrastructure is called the Knowledge Environment for Collaborative Science (KnECS).

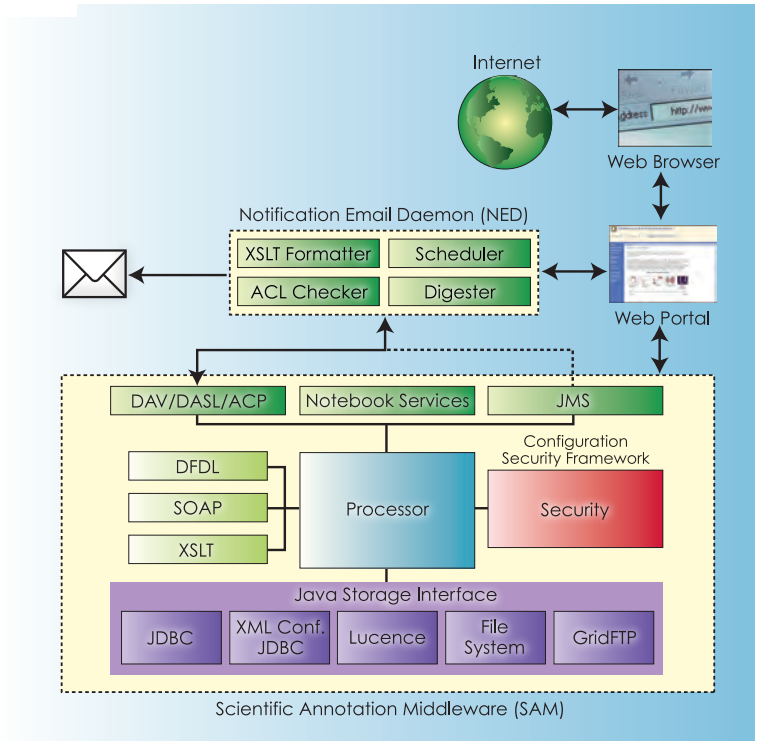

Figure 1. KnECS architecture

KnECS [2] defines a multilayer architecture, that supports a Web portal to allow users to interface with tools and collaborate with one another. The collaborative environment in KnECS is based on the CompreHensive collaborativE Framework (CHEF) [3] software, which provides support for teams and community interaction tools. KnECS also leverages Scientific Annotation Middleware (SAM) [4], which provides data and metadata services, such as searching, access control, automatic metadata extraction, data translations, and managing data provenance. KnECS provides APIs for notifications, easy access to the SAM data repository, and the construction of structured metadata, as well as templates for integrating new portlets into the Web portal.

SAND2006-2559P
April 2006

KnECS includes several knowledge tools, such as a DataBrowser portlet for advanced searching, provenance manipulation, metadata editing, and standard file operations. KnECS is available as open-source software, and considerable development has been done to make the deployment of KnECS easy.

Impact

Advanced provenance tracking, data interoperability, and multilevel application support in KnECS facilitates collaboration among scientific communities and unique facilities, allowing the formation of knowledge grids for new science areas.

Collaboratory for MS3D

The Collaboratory for MS3D (C-MS3D) [5] is building on KnECS to create a knowledge grid for the biomedical community. Researchers from UMBC, UCSF, and Sandia are creating a Web-based data sharing and tool development portal to support MS3D research [6,7], a mass-spectrometry-based approach to generating distance constraint information in macromolecules and macromolecular complexes using chemical cross-linking, proteolysis, and mass spectrometry. C-MS3D will integrate workflow tools into KnECS and add science-specific data translators and tools to provide a collaborative problem-solving environment for MS3D researchers.

Future Work

Our vision is that such knowledge grids will not only facilitate interdisciplinary collaboration, but also enable scientists to move toward tackling data-intensive and multiscale problems with systems science approaches. These approaches can enhance the discovery of knowledge gaps, clarify research priorities, and potentially accelerate scientific impact on industrial development and societal needs. The KnECS science portal and data management infrastructure provide a significant foundation creating these knowledge grids. We hope to create future science portals for other science applications, as well as continue to develop the KnECS infrastructure and add new features. Further, we are supporting the outreach and growth of our two existing collaboratories, C-MS3D and CMCS.

References

[1] The Collaboratory for Multi-scale Chemical Science Web site, http://cmcs.org/, 2006.
[2] Schuchardt, K. et al, Portal-based knowledge environment for collaborative science, to appear in a special issue of *Concurrency and Computation: Practice and Experience*.
[3] CHEF Collaborative Portal Framework Web site, http://www.chefproject.org, University of Michigan, 2006.
[4] Myers JD et al., Re-integrating the research record. *IEEE Computing in Science and Engineering*, 2003. 5(3):44—50.
[5] The Collaboratory for MS3D Web site, http://ms3d.org/, 2006.
[6] Young, M., et al., High-throughput structure determination: rapid identification of protein folds using mass spectrometry and intramolecular cross-linking. Proc Natl Acad Sci U S A, 2000. 97(11): 5802—6.
[7] Schilling, B., et al., MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. J Am Soc Mass Spectrom, 2003. 14(8): 834—50.

Acknowledgements

The CMCS project is part of the National Collaboratories Program sponsored by the U.S. Department of Energy's Office of Mathematical, Information, and Computational Sciences. Funding for C-MS3D is provided jointly by the National Institutes of Health (RR019864-01) and the National Science Foundation (Chem 0439067).

For more information, contact
Carmen Pancerella, (925) 294-3538
Christine Yang, (925) 294-2016